ANALYZING THE IMPACT OF CAR FEATURES ON PRICE AND PROFITABILITY

PROJECT INTRODUCTION

The automotive industry is undergoing rapid changes, with a shift towards electric and hybrid vehicles, increased focus on fuel efficiency, and evolving consumer preferences. As a Data Analyst, my goal is to help a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand.

APPROACH

In approaching this project, I began by utilizing Python for data analysis. After loading the dataset into a Pandas DataFrame, I performed data preprocessing, handling missing values and ensuring data quality. Exploratory data analysis was conducted to gain insights into key features, and I visualized relationships using Seaborn and Matplotlib. To understand the impact of car features on pricing, I employed a OLS regression model. I assessed the model's performance using mean squared error and visualized feature importances. Further, I extended the analysis to categorical columns, considering all factors for sensitivity analysis.

This comprehensive approach aimed to provide actionable insights for the car manufacturer, informing optimal pricing strategies and aiding in product development decisions to maximize profitability while meeting consumer demand.

DATASET OVERVIEW:

The dataset contains information about various car features, market categories, and pricing. We will leverage data analysis techniques to extract meaningful insights.

Rows - 11914

Columns - 16

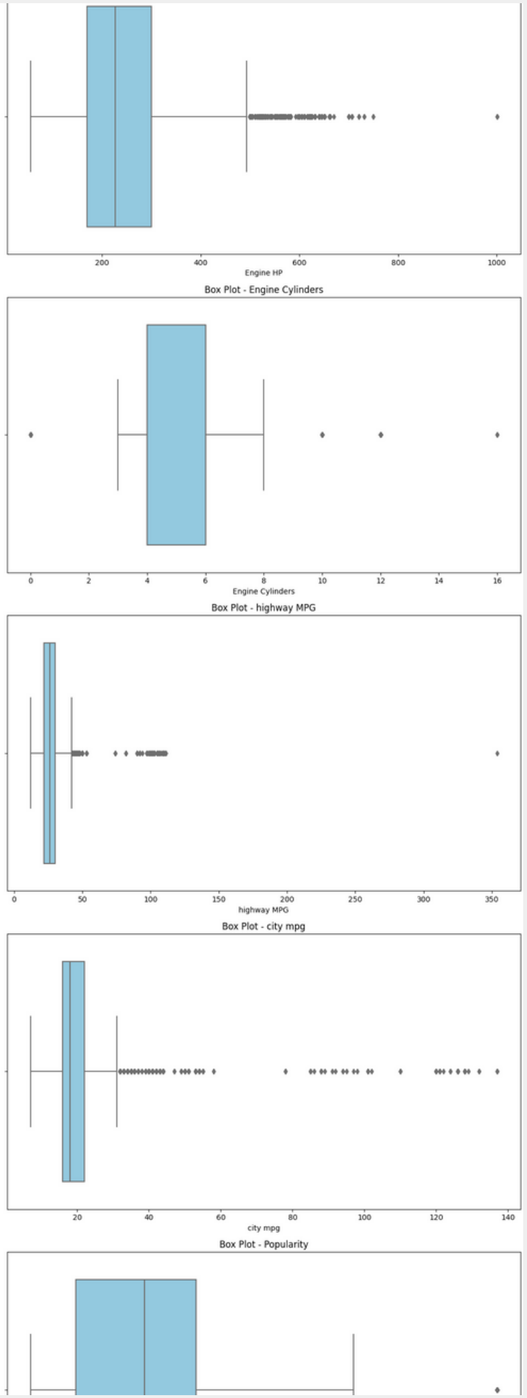
Here I am giving my github link

TECH - STACK USED:

The primary technology used in this project was ***Python***, leveraging libraries such as *Pandas for data manipulation*, *Seaborn and Matplotlib for data visualization*, and *scikit-learn for implementing machine learning models*, specifically linear regression for price prediction. Additionally, the project involved the use of popular data science tools and techniques, including **Jupyter Notebook** for code execution and documentation. The analysis also made use of statistical metrics, such as mean squared error, to evaluate the performance of the regression model. Overall, Python served as the core programming language, providing a versatile and powerful environment for the entire data analysis pipeline.

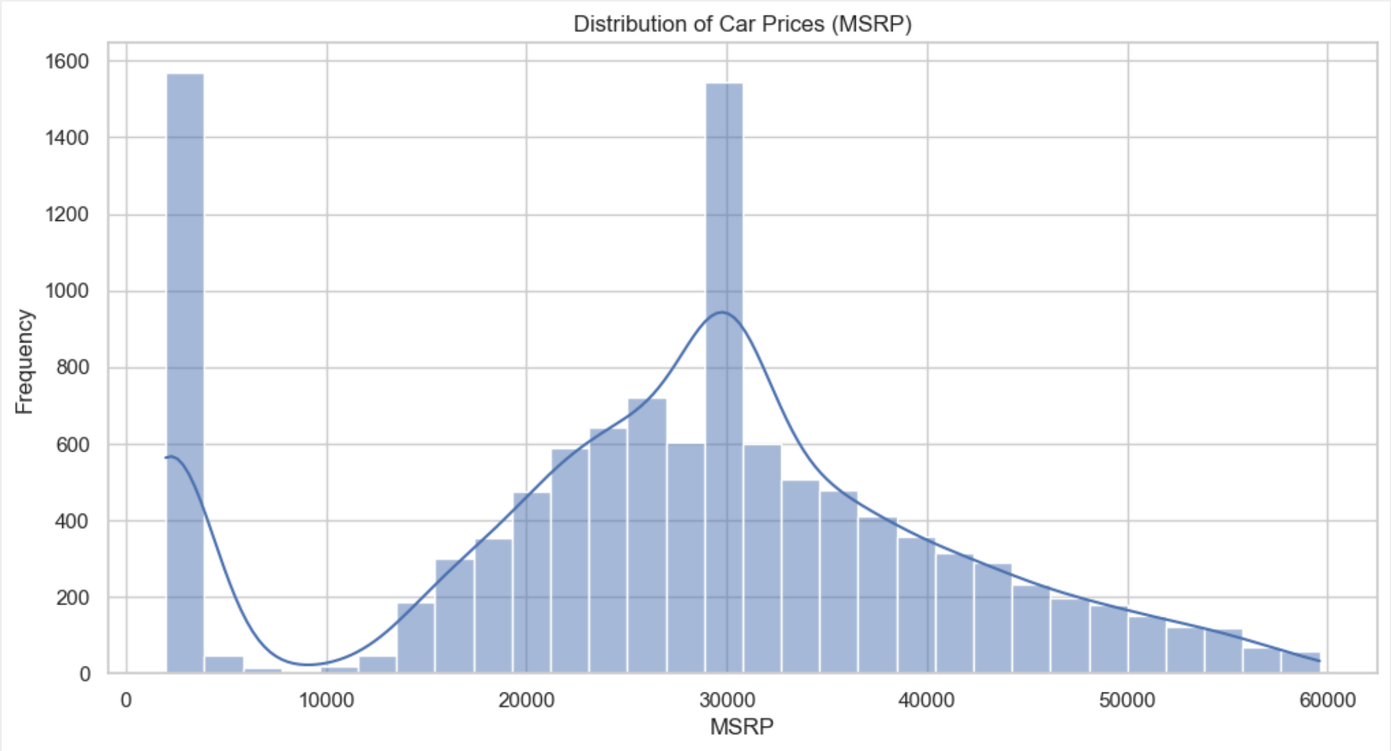
DATA CLEANING

The first step in the data analysis process was data cleaning. I began by checking for missing values and handling them appropriately, either by imputation or removal, to ensure the integrity of the dataset. I also addressed any outliers that could potentially skew the analysis. Categorical variables were encoded, and I verified the consistency and accuracy of the data types across columns. This cleaning process laid the foundation for a reliable and accurate analysis, reducing the likelihood of errors and ensuring the quality of the insights derived from the dataset.

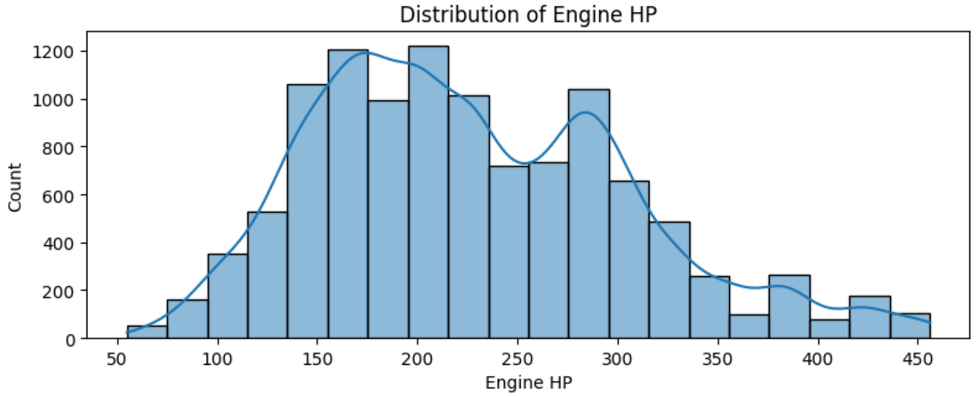


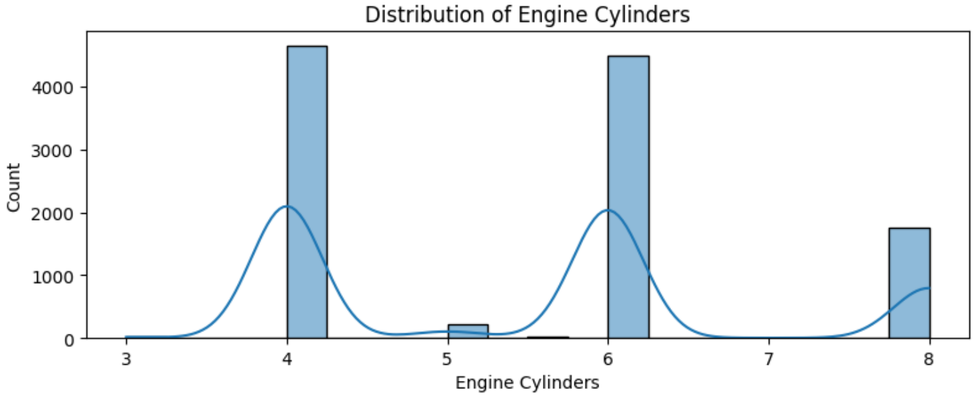
EDA FOR UNCOVERING PATTERNS AND TRENDS

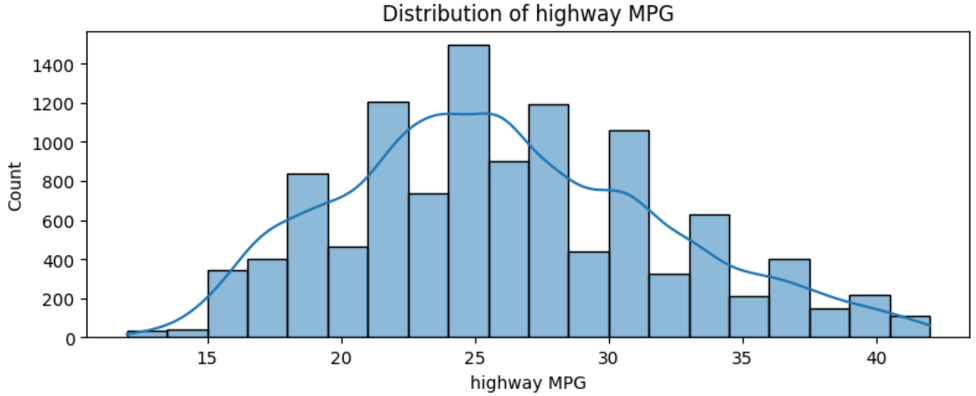
- Key Points:
  - Utilizing descriptive statistics to understand the distribution of key features.
  - Creating visualizations (histograms, scatter plots) to explore relationships.
  - Performing correlation analysis to identify potential dependencies.

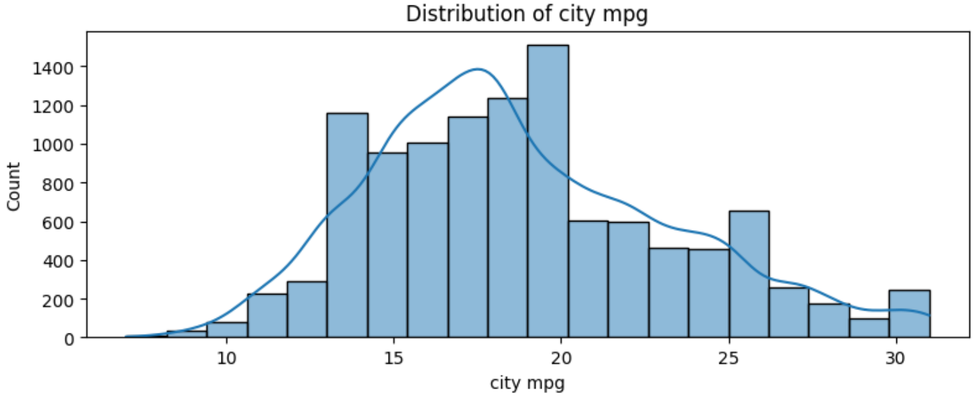


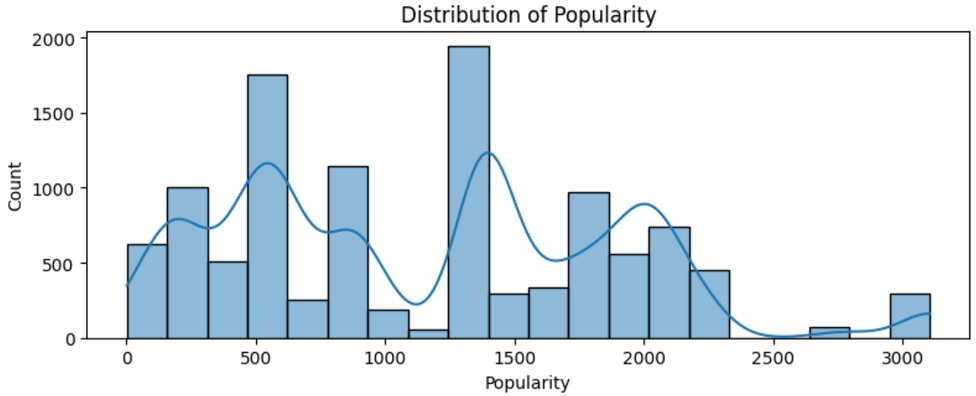
GRAPHS FOR CONTINUOUS COLUMNS

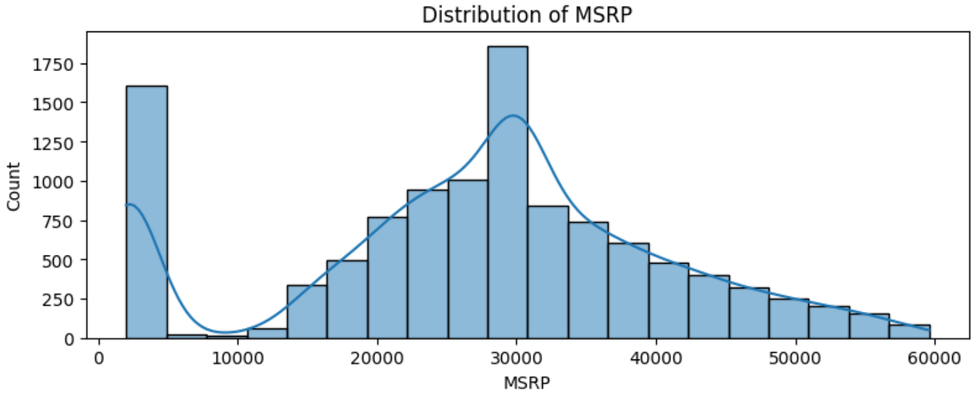




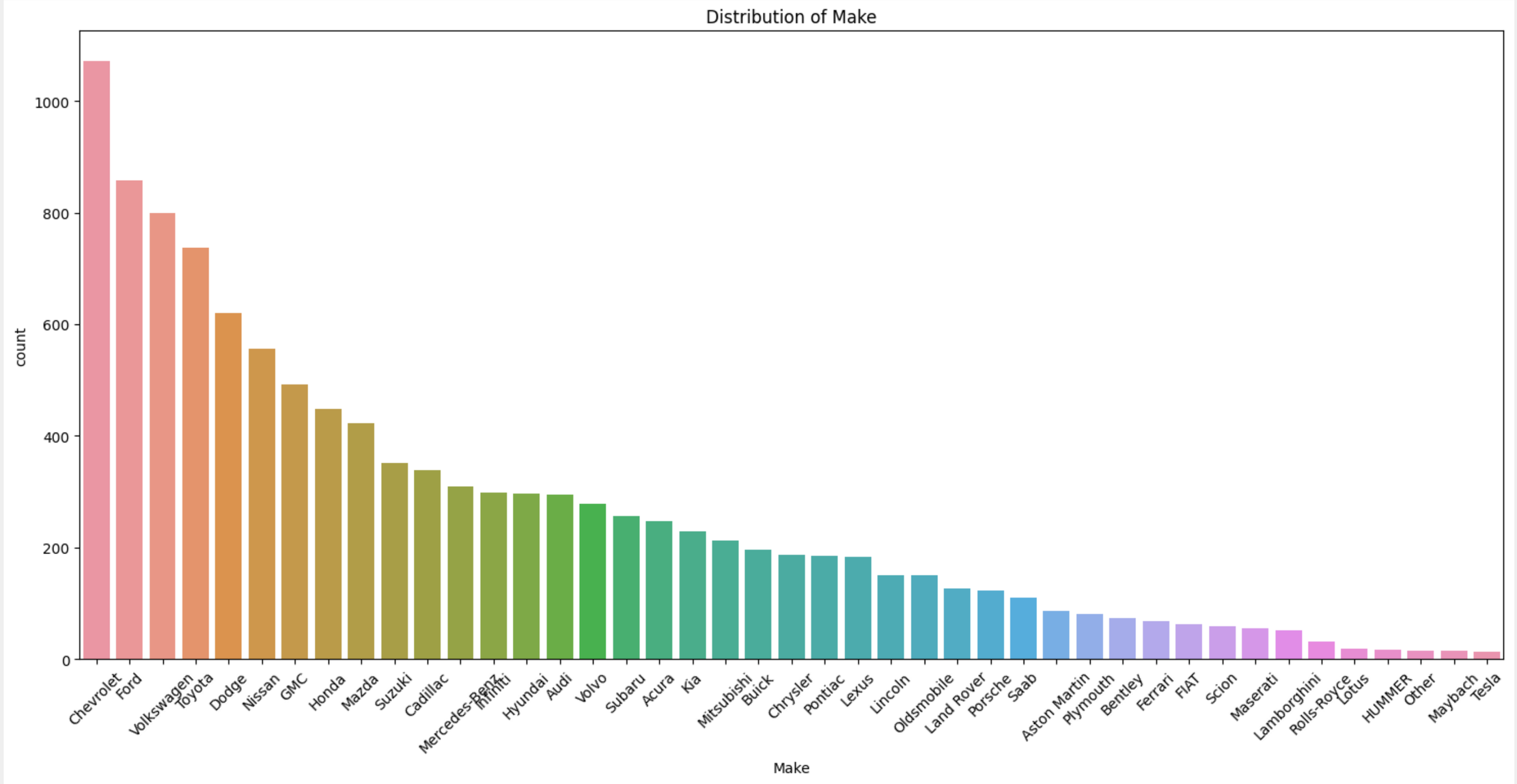


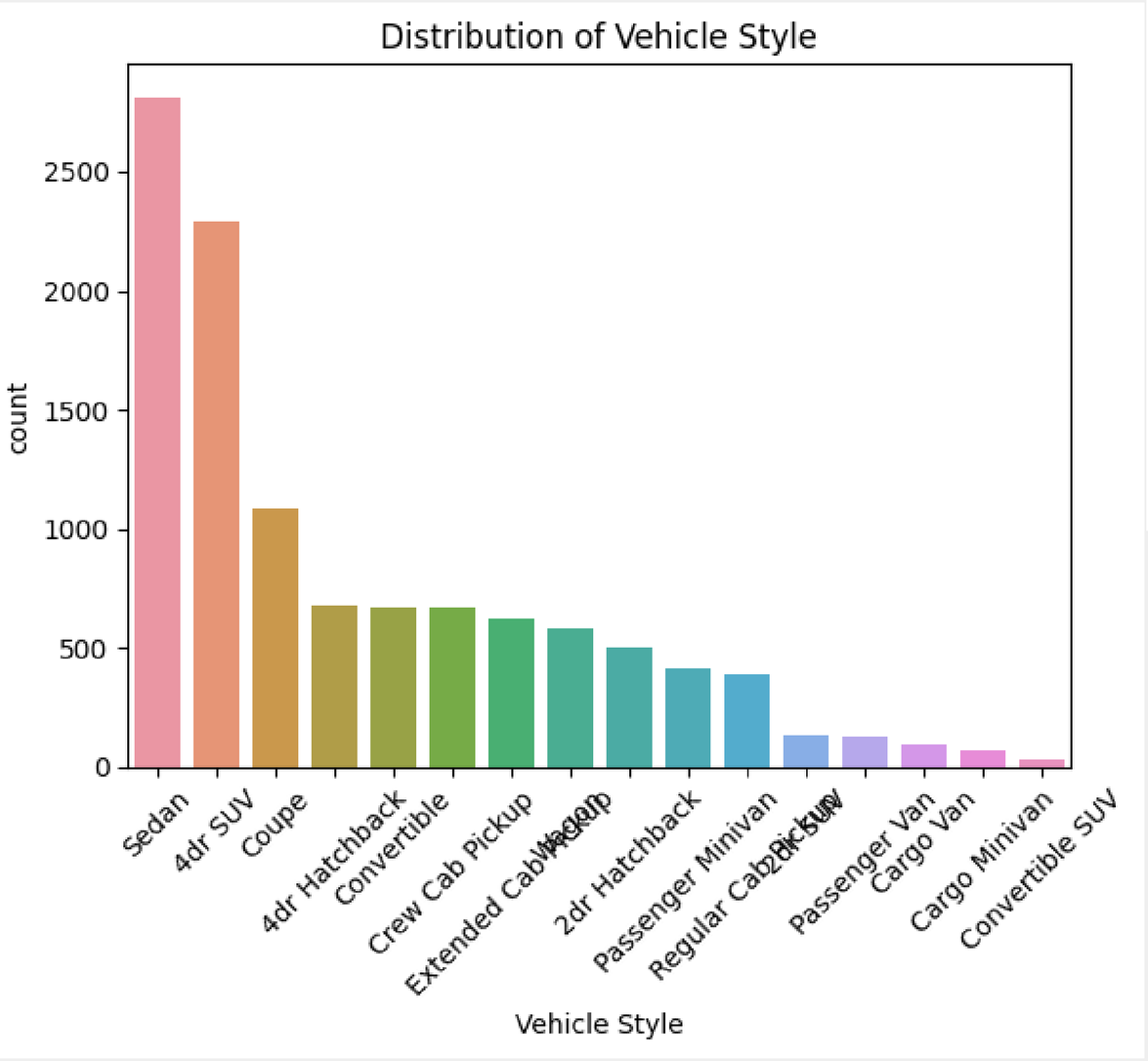


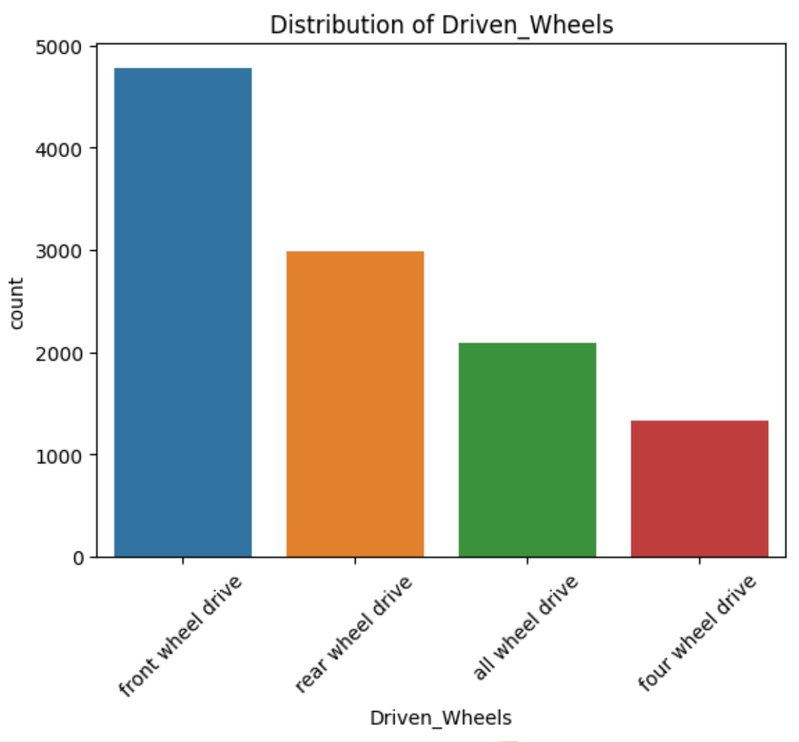


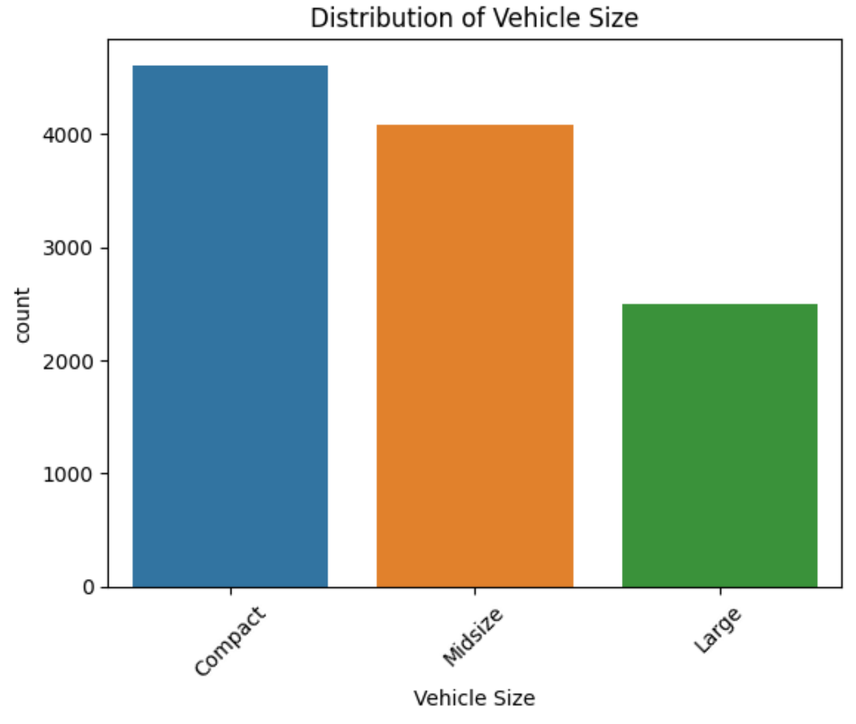


GRAPH FOR CATEGORICAL COLUMNS

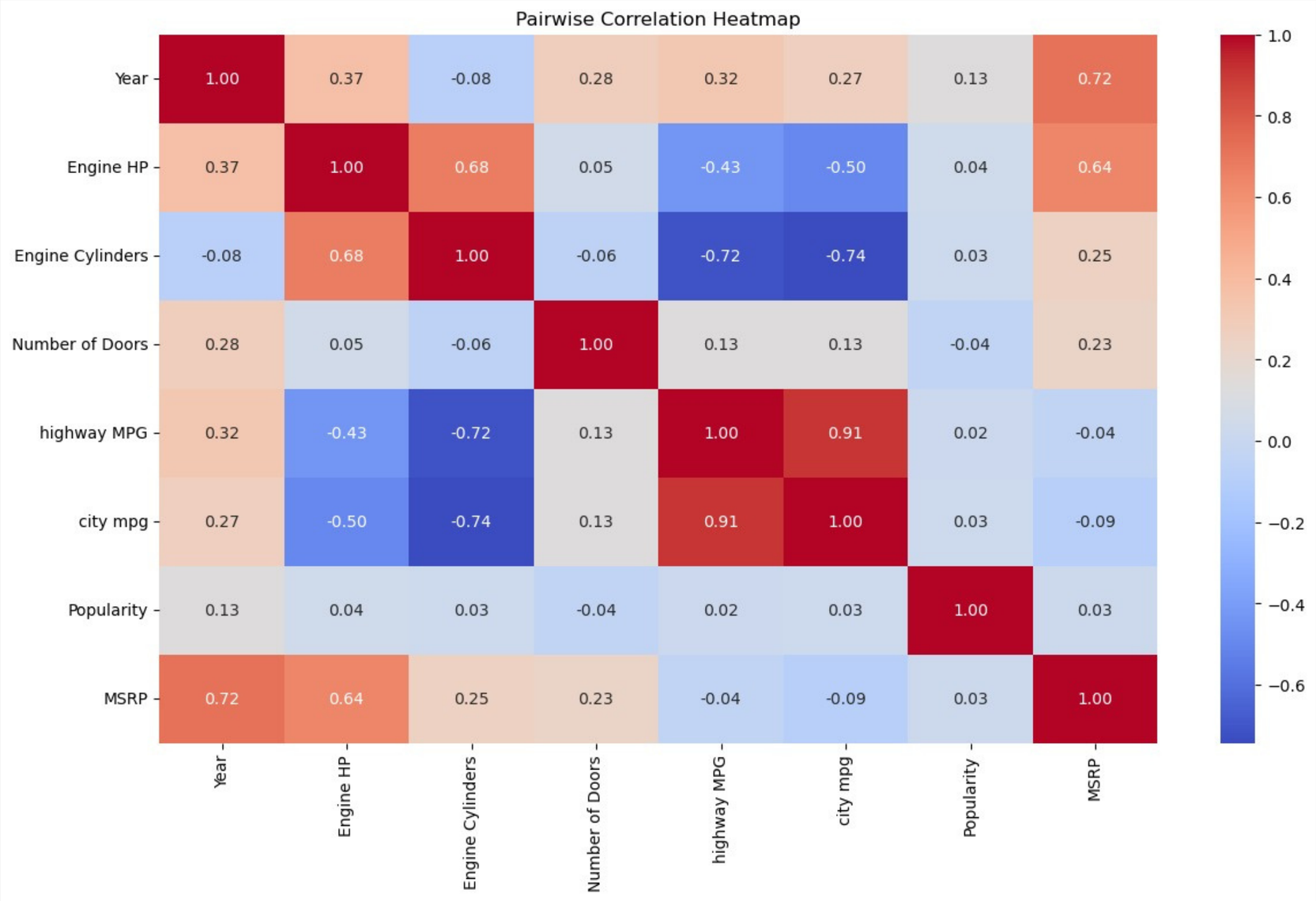



Distribution of Vehicle Style
count
2500
2000
1500
1000
500
0
Sedan
4dr SUV
Coupe
4dr Hatchback
Convertible
Crew Cab Pickup
2dr Hatchback
Passenger Minivan
Passenger Van
Cargo Van
Cargo Minivan
Convertible SUV
Vehicle Style


Distribution of Driven_Wheels
count
5000
4000
3000
2000
1000
0
front wheel drive
rear wheel drive
all wheel drive
four wheel drive
Driven_Wheels


Distribution of Vehicle Size
count
4000
3000
2000
1000
0
Compact
Midsize
Large
Vehicle Size

Heatmap



Top 5 correlated pairs:

city mpg	highway MPG	0.906756
MSRP	Year	0.720490
Engine Cylinders	Engine HP	0.677819
MSRP	Engine HP	0.640594
Engine HP	Year	0.368351

Regression Analysis

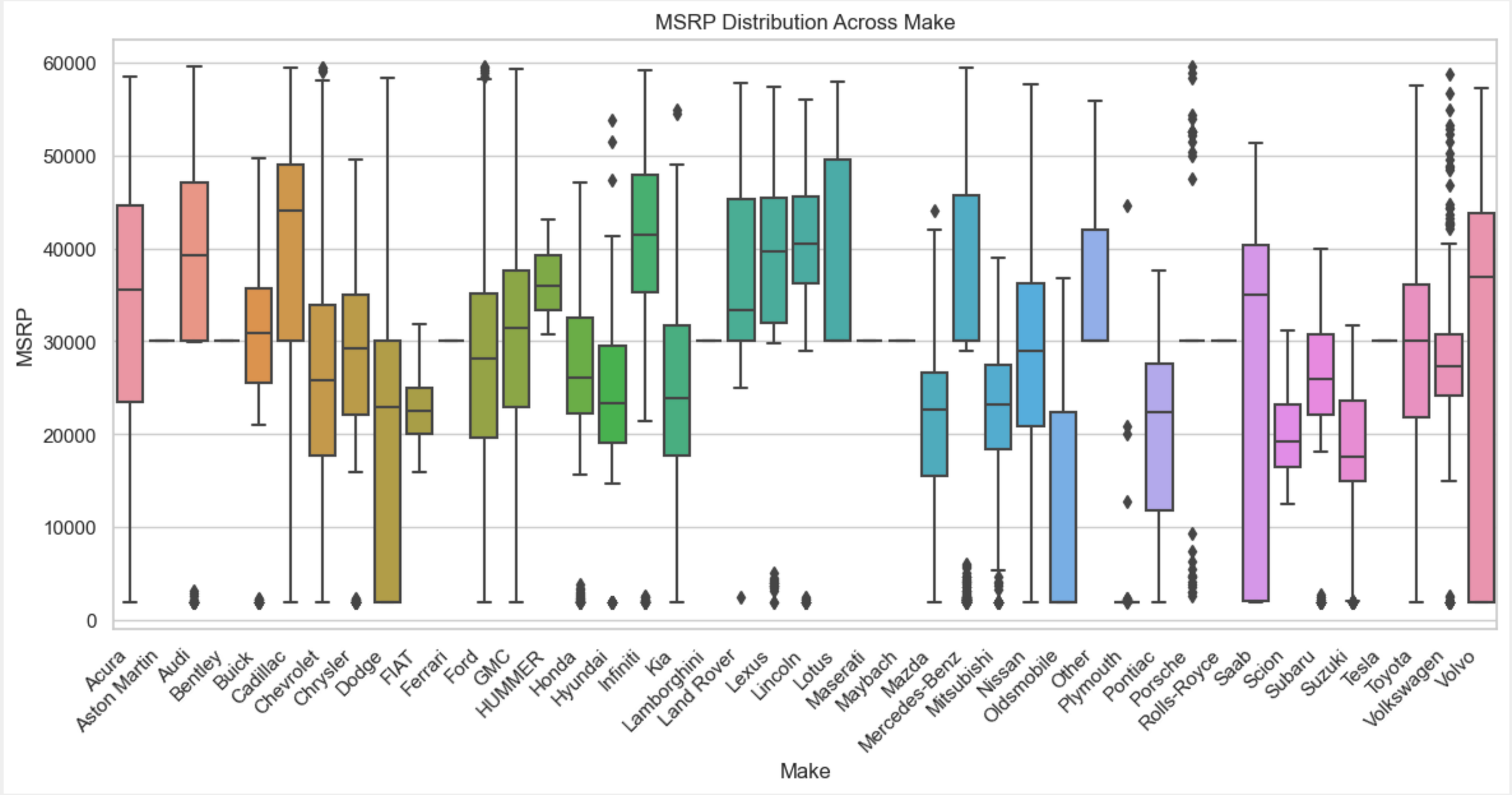
The regression analysis aims to predict the MSRP (Manufacturer's Suggested Retail Price) based on various factors. Key findings include:

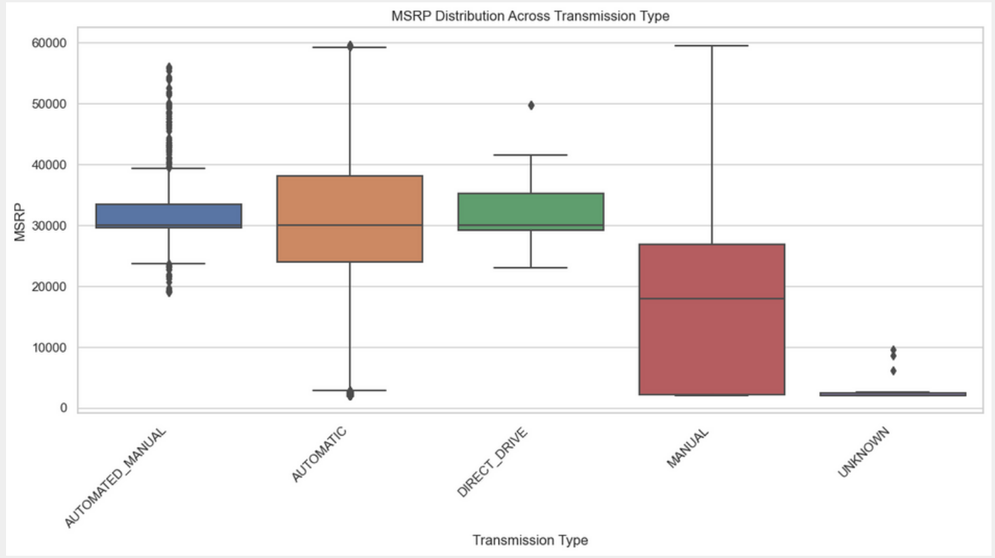
1. **Model Fit:** The model explains approximately 68.5% of the variability in MSRP (R-squared = 0.685).
2. **Overall Significance:** The model is overall significant (high F-statistic = 4858), suggesting at least one variable significantly predicts MSRP.
3. **Coefficients:**
    - 'Year' has a positive impact on MSRP, with a $1048.20 increase for each additional year.
    - 'Engine HP' also positively affects MSRP, with a $72.38 increase for each additional unit.
    - 'City mpg' negatively impacts MSRP, decreasing by approximately $100.52 for each unit increase.

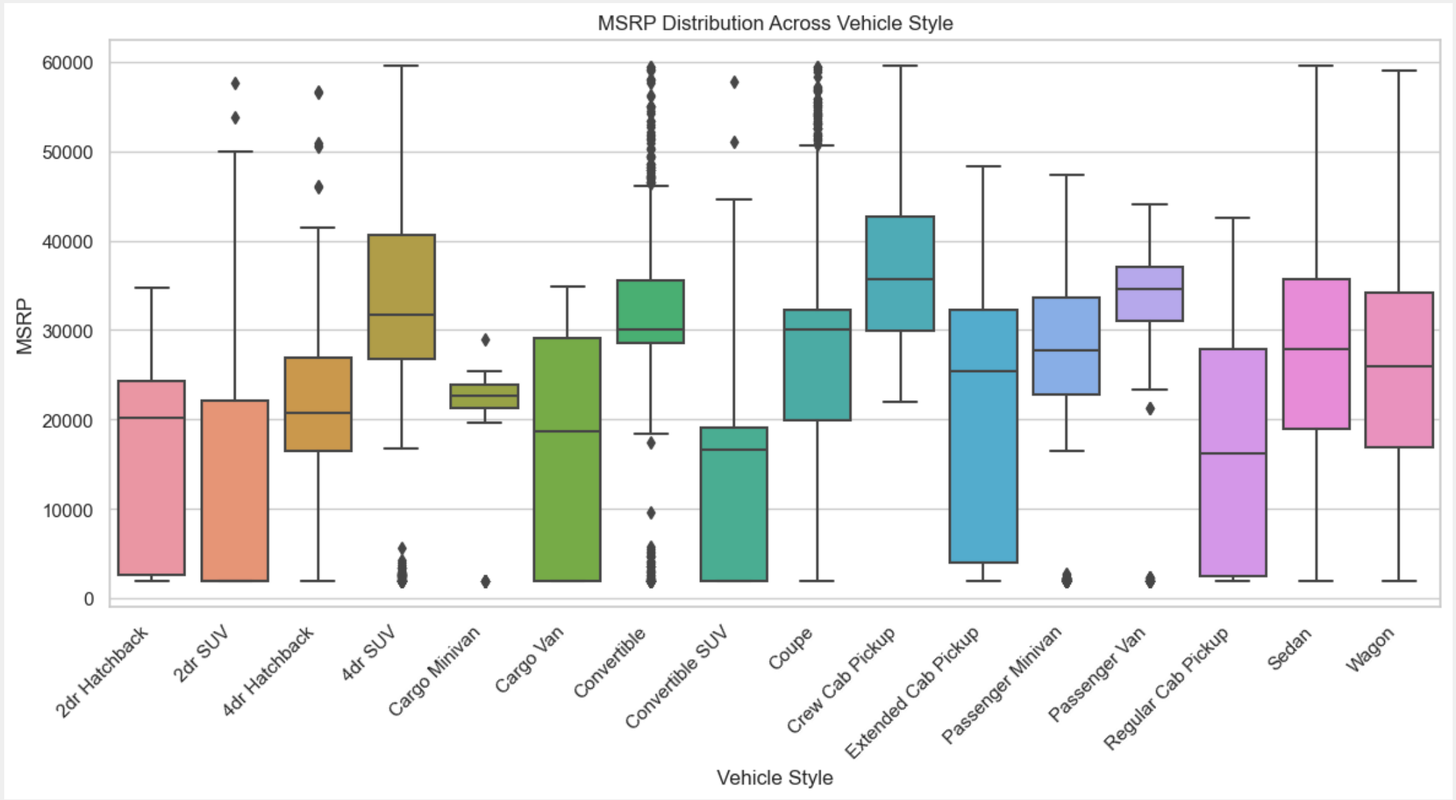
The multiple linear regression model for predicting MSRP based on the given independent variables is expressed mathematically as:

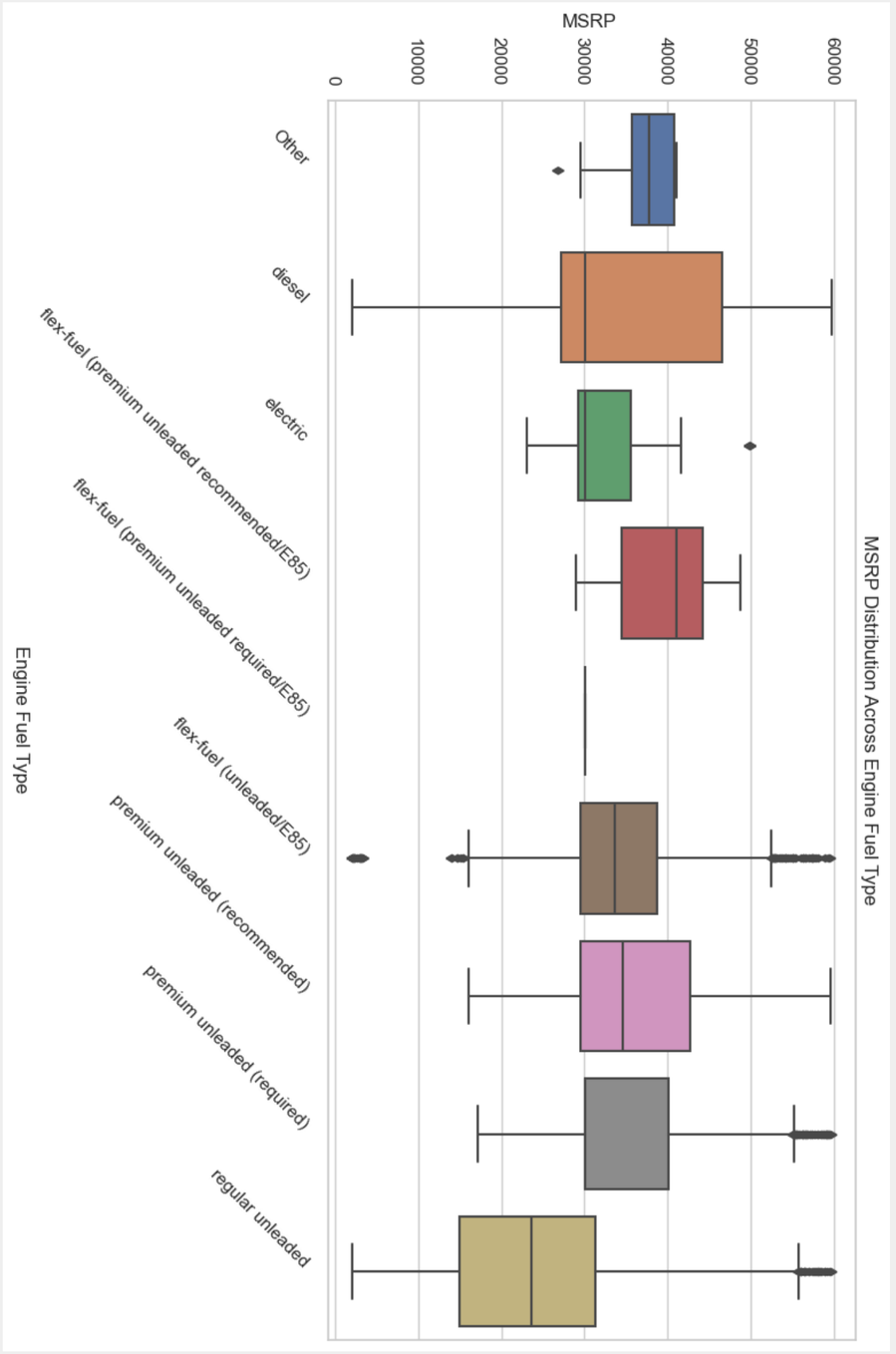
MSRP=−2.089×106−100.5202×city mpg−123.0074×highway MPG+1048.2044×Year−465.9896×Engine Cylinders+72.3753×Engine HP

Market Segmentstion




MSRP Distribution Across Transmission Type
MSRP
60000
50000
40000
30000
20000
10000
0
AUTOMATED_MANUAL
AUTOMATIC
DIRECT_DRIVE
MANUAL
UNKNOWN
Transmission Type


MSRP Distribution Across Vehicle Style
MSRP
60000
50000
40000
30000
20000
10000
0
2dr Hatchback
2dr SUV
4dr Hatchback
4dr SUV
Cargo Minivan
Cargo Van
Convertible
Convertible SUV
Coupe
Crew Cab Pickup
Extended Cab Pickup
Passenger Minivan
Passenger Van
Regular Cab Pickup
Sedan
Wagon
Vehicle Style


MSRP
0
10000
20000
30000
40000
50000
60000
Other
diesel
electric
flex-fuel (premium unleaded recommended/E85)
flex-fuel (premium unleaded required/E85)
flex-fuel (unleaded/E85)
premium unleaded (recommended)
premium unleaded (required)
regular unleaded
Engine Fuel Type
MSRP Distribution Across Engine Fuel Type

Based on the visualizations and analyses performed, here are some insights into patterns or trends in MSRP based on different segments:

Categorical Variables:

Make: MSRP varies significantly across different car makes. Some makes tend to have higher average MSRP than others.

Model: MSRP varies across different car models. Some models have higher average MSRP compared to others.

Engine Fuel Type: Different engine fuel types exhibit variations in MSRP. For some fuel types, MSRP tends to be higher on average.

Transmission Type: MSRP differs based on the type of transmission. Certain transmission types are associated with higher average MSRP.

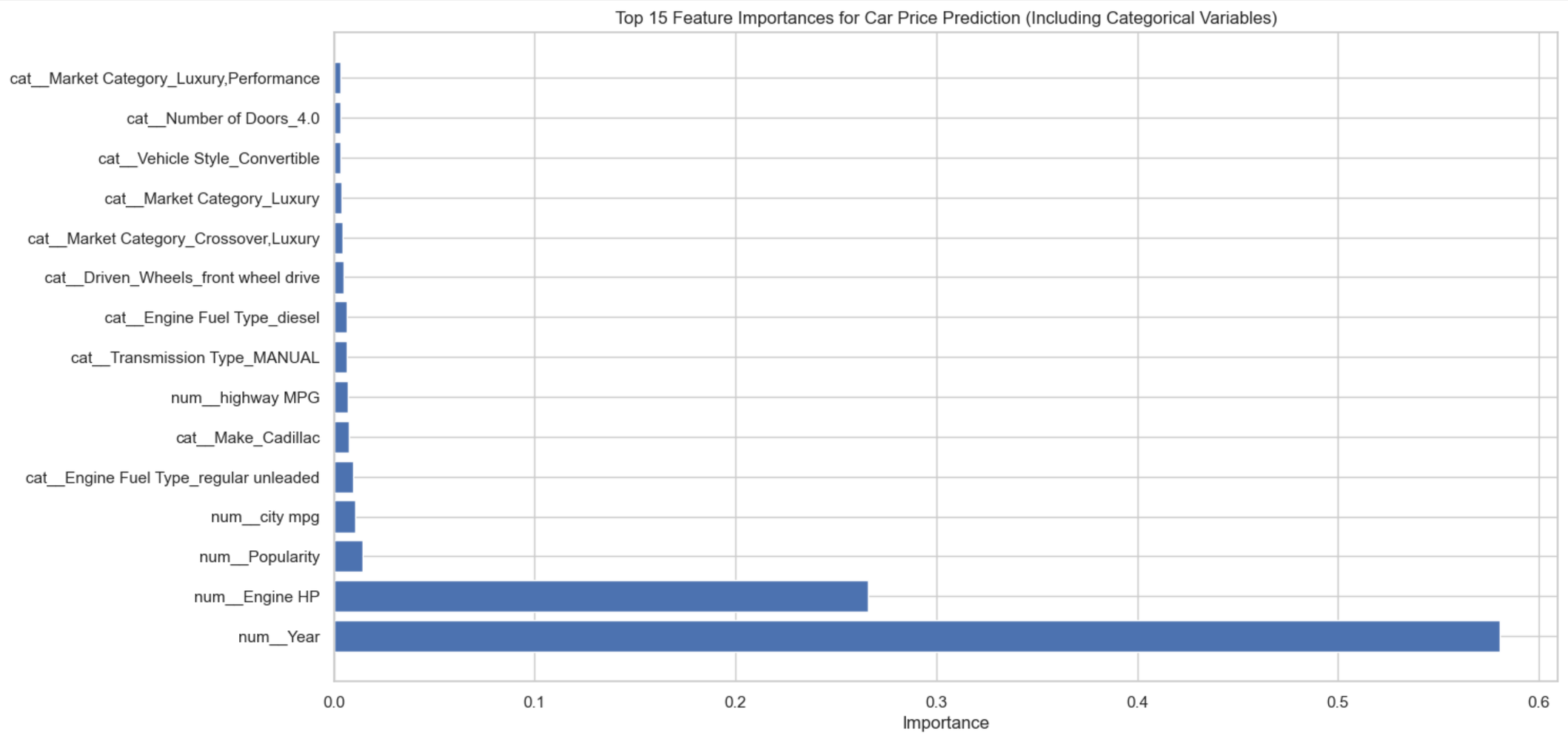
Driven Wheels: Driven wheels influence MSRP, with certain types associated with higher average prices.

Number of Doors: MSRP varies based on the number of doors, with specific counts associated with higher prices.

Market Category: Different market categories show variations in MSRP. Some market categories have higher average MSRP. Vehicle Size: Vehicle size impacts MSRP, with larger sizes often associated with higher prices.

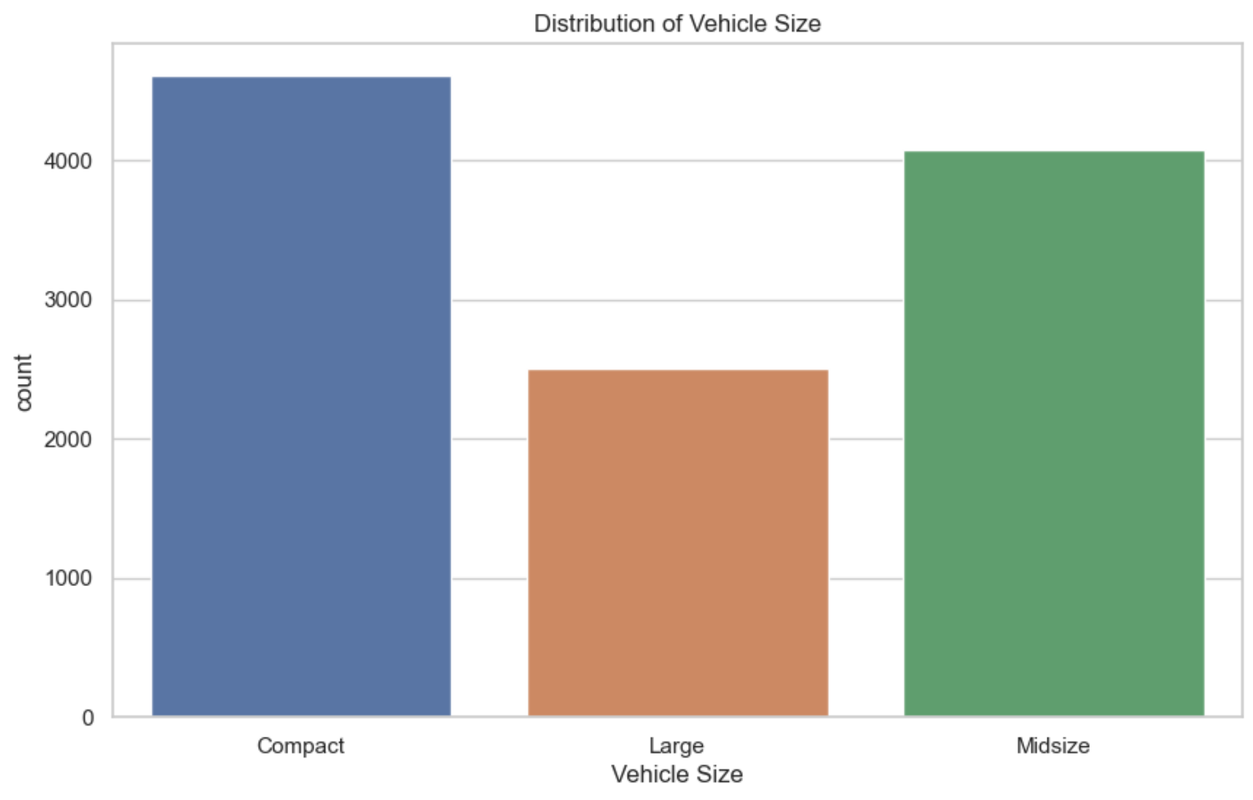
Vehicle Style: Vehicle styles exhibit variations in MSRP. Certain styles tend to have higher average MSRP.

Overall Trends:
Luxury car makes and models often have higher MSRP. Performance-related features (higher Engine HP, more Cylinders) are associated with higher MSRP. Larger and more luxurious vehicle styles tend to have higher MSRP. These observations provide a general overview, and further analysis or specific inquiries into particular segments can reveal more detailed insights.

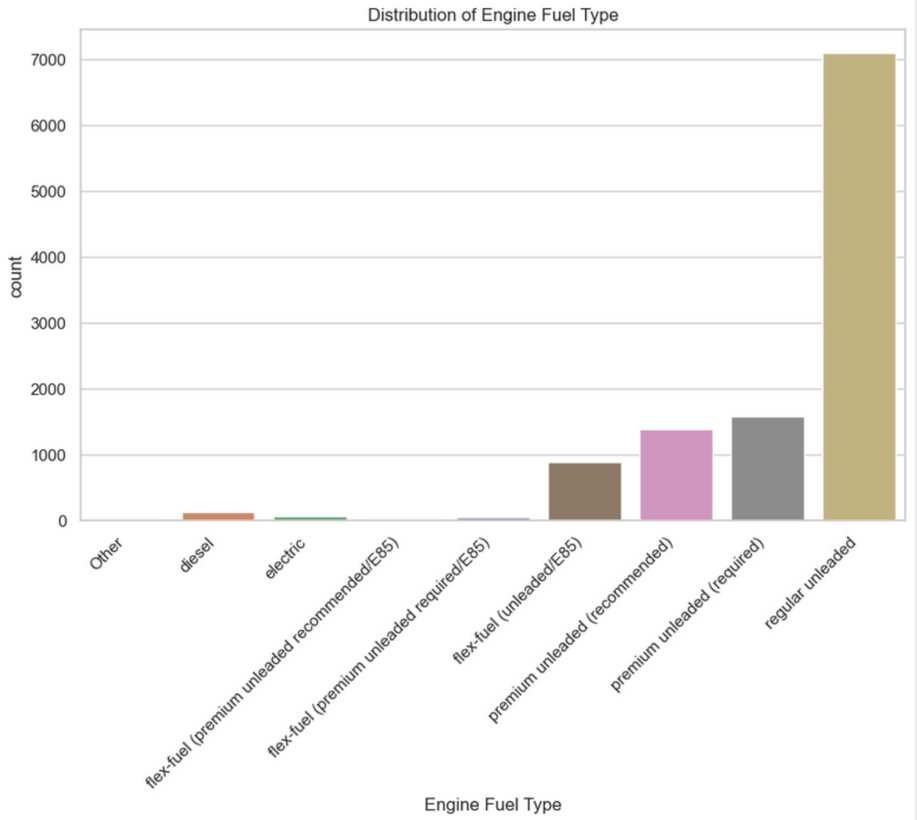


From this graph we can interpret that Year and Engine HP has maximum effect on car MSRP

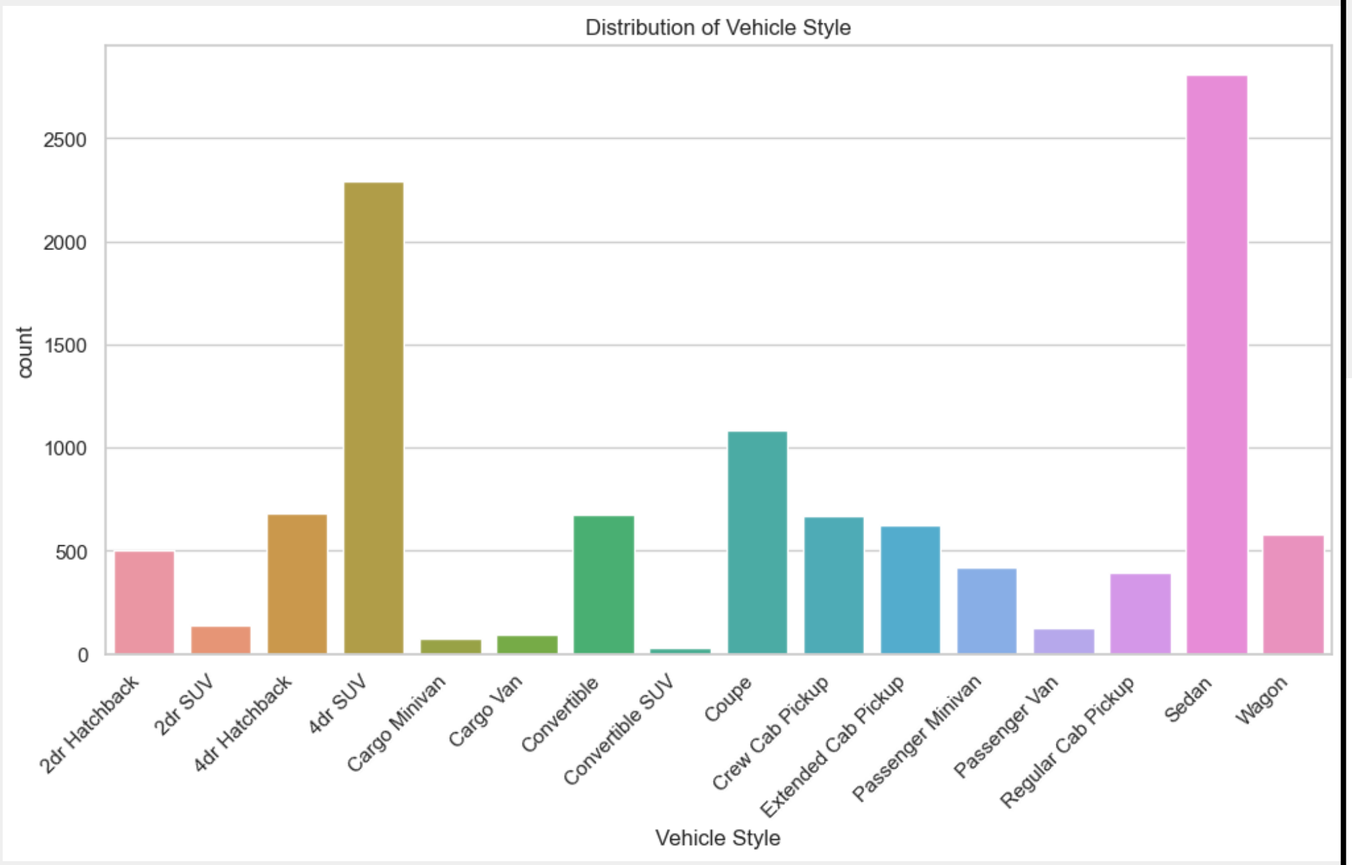
CONSUMER PREFERENCES:



Customer prefers compact and Midsize car more

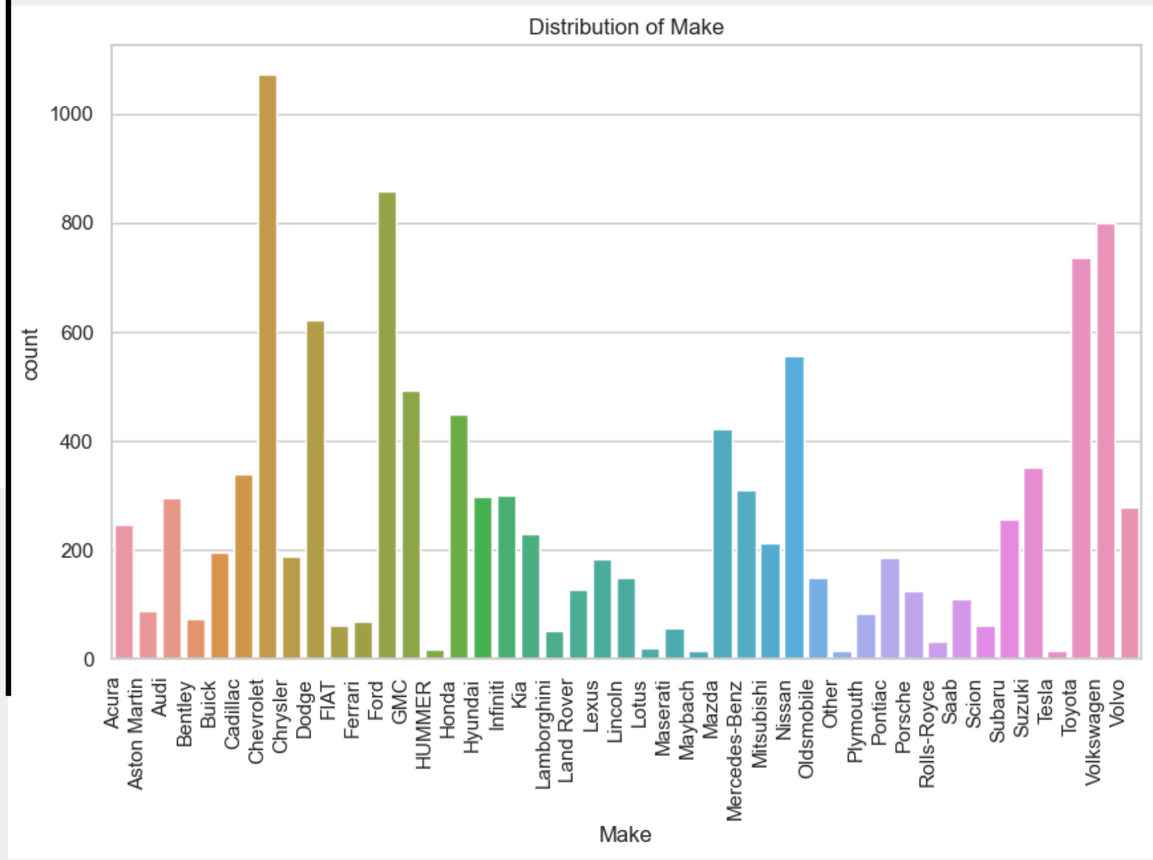


Customer prefers regular unleaded fuel type cars more

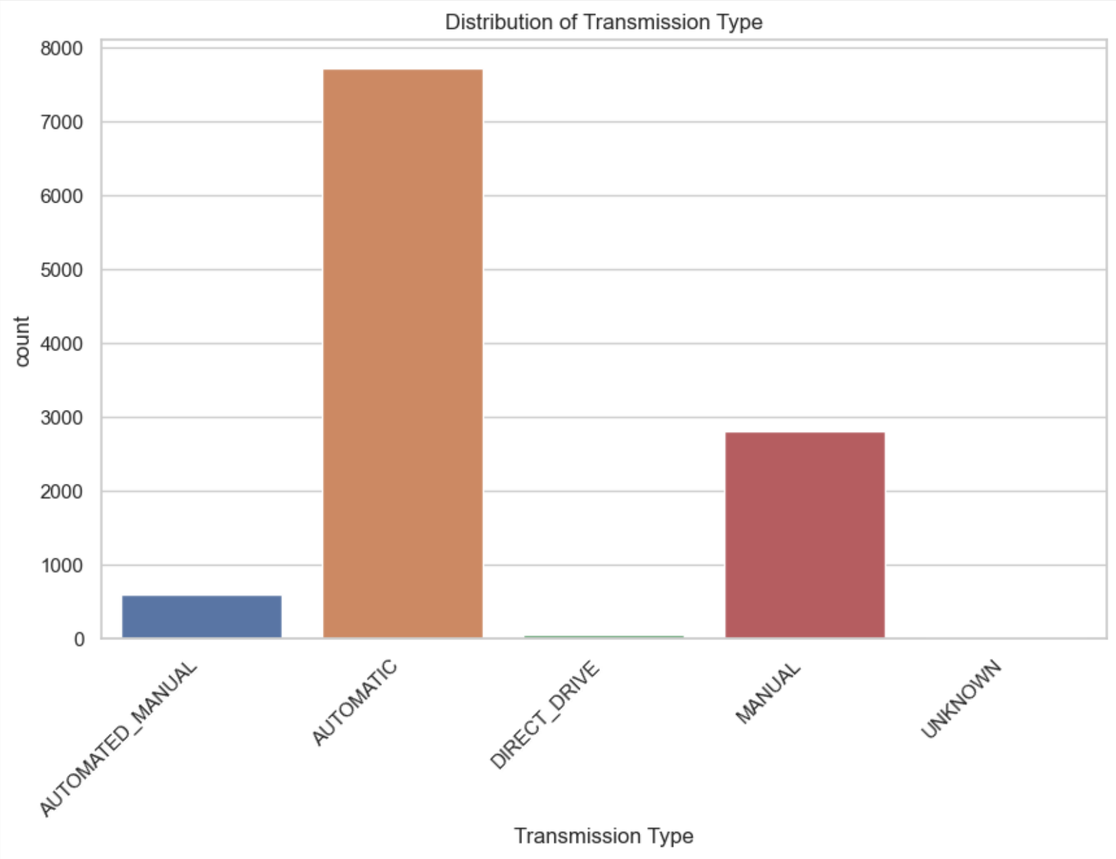


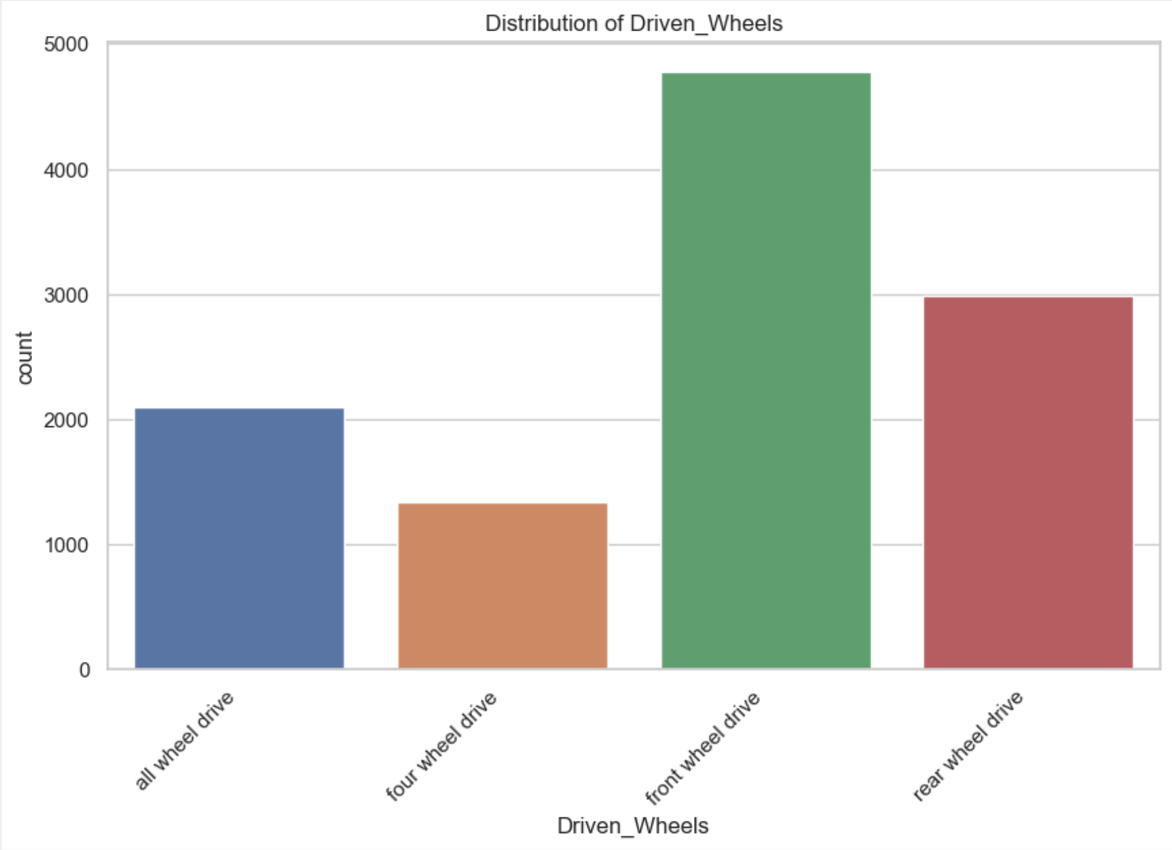
Customer prefers Sedan and 4dr Hatchback style cars more

Customer prefers Chevrolet Ford, Volkswagen and Toyota manufacturer car more



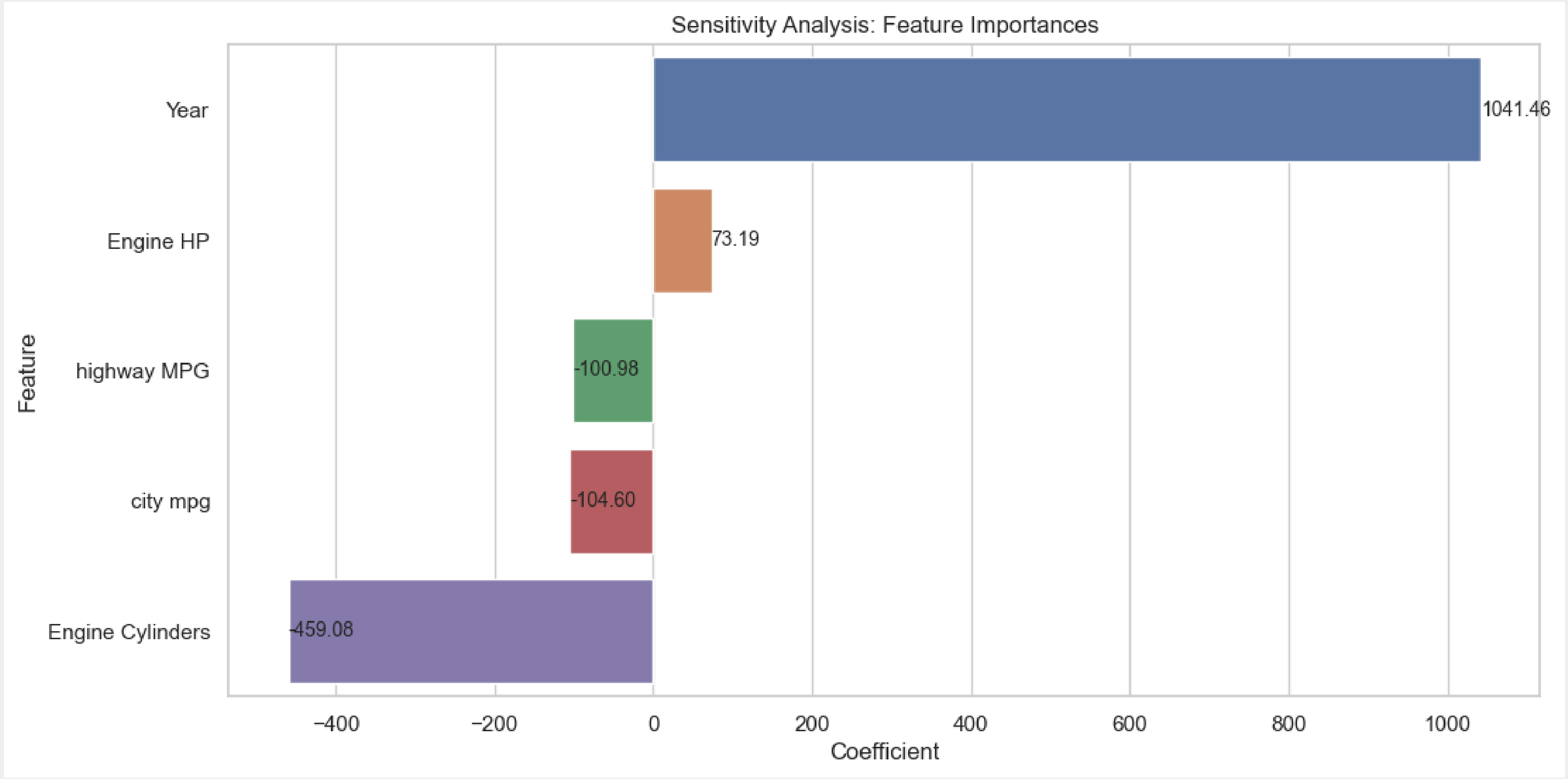
Customer prefers Automatic cars more





Customer prefers Front Wheel drive cars more

Sensitivity Analysis



Our analysis reveals that certain car features exhibit distinct impacts on the Manufacturer's Suggested Retail Price (MSRP). Specifically, a more recent manufacturing year and increased engine horsepower contribute positively to the MSRP. On the other hand, enhanced fuel efficiency, reflected in higher highway and city miles per gallon (MPG), as well as a lower number of engine cylinders, is associated with a decrease in the MSRP. These insights are crucial for car manufacturers to make informed decisions regarding pricing and product development, ensuring alignment with consumer preferences and maximizing profitability.

Conclusion:

The project concludes with valuable insights for the car manufacturer to optimize pricing and product development decisions. Key findings and recommendations include:

Important Features Affecting MSRP:

'Engine HP' and 'Year' emerged as crucial factors influencing car prices. Manufacturers should focus on enhancing these aspects in product development to potentially increase profitability.

Market Segmentation:

Analysis of market segments ('Make') provides a deeper understanding of consumer preferences. Tailoring strategies based on these segments can help the manufacturer meet specific demands.

Sensitivity Analysis:

Numeric features like 'Engine HP' and 'Year' significantly impact pricing. Recognizing their importance, the manufacturer can strategically incorporate these features to appeal to consumer preferences.

Recommendations:

Emphasize technological advancements ('Engine HP') and consider the model year ('Year') in product development.
Implement targeted pricing strategies based on market segments ('Make').

Optimizing Competitiveness:

By aligning product features with consumer preferences and adopting strategic pricing, the manufacturer can enhance competitiveness in the dynamic automotive market.
In summary, the project equips the manufacturer with actionable insights to strike a balance between meeting consumer demand and maximizing profitability, ensuring a sustainable and competitive position in the automotive industry.

SUMMARY

The project involved a detailed analysis of a car dataset to assist a car manufacturer in optimizing pricing and product development decisions. Through data exploration, sensitivity analysis, and market segment analysis, key features impacting car prices and consumer preferences were identified. The analysis extended to both numeric and categorical features, including market segments based on car manufacturers. Strategies were proposed for the manufacturer to balance consumer demand and profitability, leveraging insights from the data. The project utilized Python for analysis, and visualizations were created to effectively communicate findings to stakeholders.

THANK YOU !